INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Comparing /b/ and /d/ with a single physical model of the human vocal tract to visualize droplets produced while speaking

Takayuki Arai[1], *Tsukasa Yoshinaga*[2], *and Akiyoshi Iida*[3]

[1]Sophia Univ. (Japan), [2]Osaka Univ. (Japan), [3]Toyohashi Univ. of Technology (Japan)
arai@sophia.ac.jp

Abstract

The BMW-RL model, a physical model of the human vocal tract that produces /b/, /m/, /w/, /r/, and /l/, has been utilized to investigate not only each single sound but also consonant clusters such as /br/. This model started gaining attention in 2021 in light of the COVID-19 pandemic, as it can demonstrate how droplets are expelled from the lips when producing the /b/ sound by applying a laser sheet. In this study, we redesigned the model to produce both /b/ and /d/, since both are voiced plosives and the only difference is the place of articulation. With the original BMW-RL model, the first half of the tongue rotates and produces /r/ and /l/, while in the newly proposed model, the width of the tongue is wide enough to make a complete closure at the alveolar position for producing /d/. We tested how the different place of articulation affects the ways of expelling droplets by using the single model producing /b/ and /d/ and found that more droplets were expelled with /b/ than /d/.

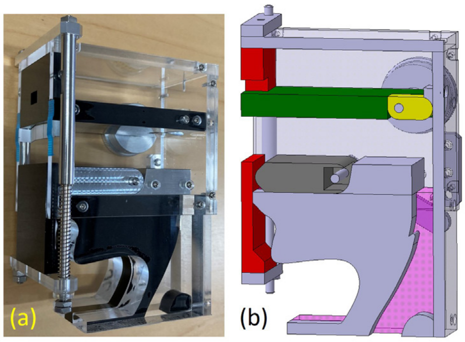
Index Terms: COVID-19, visualizing droplet cloud, vocal-tract models, plosive sounds

1. Introduction

The BMW-RL model, a physical model of the human vocal tract that produces /b/, /m/, /w/, /r/ and /l/ sounds, demonstrated that a single model could be used to produce different consonants and vowels [1]. Since this model can produce /b/ and /r/, it has been used to demonstrate how the consonant cluster /br/ is produced. While these models were originally developed for pedagogical purposes in acoustic phonetics and speech science, they are now being applied for pronunciation training, such as practicing the English /r/ and /br/ sounds for native Japanese speakers.

Since the advent of the COVID-19 pandemic, many studies have investigated the role of human speech production in expelling droplets and air flow with aerosol (e.g., [2]). To visualize droplet clouds and aerosol during speech production, a laser sheet is commonly used. However, as lasers carry a certain risk for human health, [3] used physical models of the human vocal tract to avoid such risks. In [3], two types of measurement were utilized: 1) visualizing exhaled breath during vowel production by using a head-shaped model with a lung model, and 2) visualizing droplet clouds when producing the voiced bilabial plosive sound by using two types of articulatory physical models, one of which is the BMW-RL model, with and without a lip model. The results showed that 1) the head-shaped and lung models could visualize how small particles scatter during vowel production in different conditions, and 2) the vocal-tract models could visualize how droplets are expelled with the /b/ sound in different conditions.

Figure 1: *BMW-DN model.*



As an extension of the previous BMW-RL model [2], we have newly designed and developed the BMW-DN model, which can produce /b/, /m/, /w/, /d/, and /n/ sounds. This new model enables us to compare /b/ and /d/ sounds in terms of how droplets are expelled.

2. Proposed Model

Figure 1 shows the proposed model (a) and its schematic illustration (b) created by cutting the model along the midsagittal plane (only the right portions of the model are drawn). The design of this model is based on the BMW-RL model [1]. The BMW-DN model has an oral cavity and a nasal cavity, which is located on top of the oral cavity. Between the two cavities, there is a velopharyngeal port, and the port's opening/closing is controlled by a knob with a rotation angle that determines the degree of velopharyngeal coupling. The rotating piece is 10 mm wide × 10 mm high × 15 mm long. We can also open and close the lips by raising the lower lip. With the open lips and the closed velopharyngeal port, the model produces the /a/ vowel sound, since the oral cavity is widely open and there is a narrow constriction in the pharyngeal cavity. The dimensions of the cross-section of the oral cavity and the nasal cavity are both 45 mm × 20 mm. The nasal cavity is 75 mm long and has a front-end block with a single nostril that is 10 mm wide × 6 mm high × 10 mm deep.

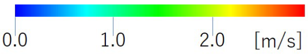
3. Measurements

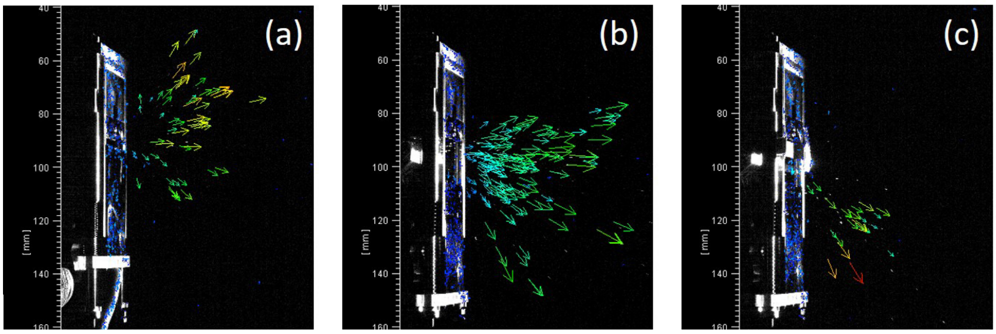
For the measurements, the BMW-DN model was used with and without simulated teeth, which were approximately 7 mm in height and constructed of a thick hard paper.

3.1. Method

The droplets are visualized by using a "simulated saliva" solution [4] made from pure glycerin (76 g) and distilled water (1 L). Approximately 30 μL of the simulated saliva solution is placed on the top surface of the lower lip for /b/ and on the top

Figure 2: *Snapshots of the results based on the PIV computation with the BMW-DN model.*
(a) /b/ sound.
(b) /d/ sound with no teeth.
(c) /d/ sound with the simulated upper teeth.





surface of the tongue tip for /d/. After positioning the saliva solution, each articulator is raised to form closed gestures. During video and audio recording, a light sheet is passed over the midsagittal plane of the models. The light sheet was produced by combining an LED illuminator (HARDsoft Microprocessor Systems, IL-106G) and projection lens (PL-180-VIS). The vocal-tract model is attached to the house of the reed-type sound source, which is connected to an air pump. When the air pump is pressed, the air stream moves through the reed-type sound source and produces a glottal sound. At the same time, the pressure builds up inside the oral cavity. When the closure is released with the articulator's opening gesture, a bilabial plosive sound or an alveolar plosive sound followed by a vowel sound is produced. The video recordings were done using a high-speed camera (Vision Research, Phantom T1340 72GB) at 3273 frames/s. The audio recordings were done using a sound level meter (Rion, NA-28) and an audio interface (RME, Fireface UC) with a sampling frequency of 48 kHz. The microphone was placed approximately 270 mm away from the mouth of the model (with 45 degrees above the horizontal).

The measurements were done in three different ways. The first measurement was for the /b/ sound, while the second and third were for the /d/ sound. In the second measurement, the model had no teeth, while in the third measurement, the simulated upper teeth were placed at the anterior part of the oral cavity of the model.

3.2. Results

We calculated particle velocities by applying a particle image velocimetry (PIV) technique with the program provided by Seika Digital Image (the color of each arrow corresponds the speed in the colormap). Figure 2 shows snapshots of the results with the BMW-DN model for each measurement. In all cases, the maximum A-weighted sound pressure level was approximately 97–99 dB. As we can see in Fig. 2, the droplets were clearly expelled when the /b/ sound was produced. On the other hand, fewer droplets were produced when the /d/ sound was produced with the simulated teeth. This is presumably because the place of articulation of /d/ is farther back than /b/ and the upper teeth blocked the trajectories of the droplets. As a result, the directions of the droplets were limited at a certain range of degrees.

4. Discussion and conclusions

In this study, we designed a new physical model of the human vocal tract based on the BMW-RL model [1]. The newly proposed model, BMW-DN, can produce not only vowel /a/ but also several consonants, including /b/, /m/, /w/, /d/, and /n/.

The original BMW-RL model was used to visualize droplets produced when articulating the /b/ sound [3]. Since the proposed BMW-DN model can produce both voiced plosive sounds, /b/ and /d/, of which the only difference is the place of articulation, it is better suited to compare both sounds simulating a single speaker producing both sounds. Conducting measurements with the laser against actual human beings comes with the risk of damage to human health, so performing the measurements with physical models is a sensible way to overcome this problem. Our experimental results with the proposed BMW-DN model showed that more droplets were expelled when the bilabial sound /b/ was produced than the alveolar sound /d/, as expected. Especially, it was due to the teeth block droplets from being expelled at certain degrees and limit the angle of their trajectories. Through the observations of the measurements, the sizes of droplets vary as well as their counts among the measurements. Especially, it is likely that more droplets are produced with smaller sizes at the teeth than the situation without teeth. Furthermore, by focusing on some droplets with the rapid movement, it is estimated that the particular ones moved with approximately 3 m/s in speed (red arrows in Fig. 2). This result shows us a risk of spreading viruses while speaking. It is also reported that healthy people produce droplets between the size of 0.1 and 10 μm [5]. In the future, we would like to measure the characteristics of droplets, including sizes and counts produced by the proposed model in different conditions.

5. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers 21K02889 and Sophia University Special Grant for Academic Research (Research in Priority Areas).

6. References

[1] T. Arai, "Integrated mechanical model for [r]-[l] and [b]-[m]-[w] producing consonant cluster [br]," in *Proc. of INTERSPEECH*, pp. 979–983, 2017.
[2] K. Onishi, A. Iida, A., M. Yamakawa and M. Tsubokura, "Numerical analysis of the efficiency of face masks for preventing droplet airborne infections," *Physics of Fluids*, vol. 34, no. 3, 033309, 2022.
[3] T. Arai, "Vocal-tract models to visualize the airstream of human breath and droplets while producing speech," in *Proc. of INTERSPEECH*, pp. 3171–3175, 2021.
[4] M. P. Wan, C. Y. H. Chao, Y. D. Ng, G. N. Sze To, and W. C. Yu, "Dispersion of expiratory droplets in a general hospital ward with ceiling mixing type mechanical ventilation system," *Aerosol Sci. Technol.*, vol. 41, no. 3, pp. 244–258, 2007.
[5] H. Zhang, D. Li, L. Xie, and Y. Xiao, "Documentary research of human respiratory droplet characteristics," *Procedia Engineering*, vol. 121, pp. 1365-1374, 2015.